



PROYECTO INTEGRAL: Planificación urbana y gestión de la movilidad

CURSO DE ESPECIALIZACIÓN EN INTELIGENCIA ARTIFICIAL Y BIG DATA

CURSO 2024-2025

Los organismos públicos están apostando por introducir el eGovern (gobierno electrónico) en sus procesos de planificación urbana y gestión de la movilidad. El eGovern hace referencia a la transformación digital de los procesos de negocio de los gobiernos, es decir, el uso de soluciones software como soporte a la ejecución de las actividades y servicios que desempeñan los organismos públicos.

Estas soluciones pueden resultar aún más efectivas al integrar tecnologías de Inteligencia Artificial y Big Data, que permiten anticipar comportamientos, generar alertas inteligentes y apoyar la toma de decisiones en tiempo real.

La contaminación atmosférica en las ciudades representa un desafío clave para la salud pública y la sostenibilidad. La capacidad de predecir, detectar y actuar ante episodios de alta contaminación es esencial para mejorar la calidad de vida y optimizar la movilidad urbana.

En el tramo final del Ciclo de Especialización en Inteligencia Artificial y Big Data del IES Dr. Lluís Simarro Lacabra de Xàtiva, debes desarrollar como alumno una solución que integre todos los módulos del ciclo ambientada en este caso de uso realista, transversal y alineado con desafíos urbanos actuales.

Guía del proyecto integrador – Curso de especialización en Inteligencia Artificial y Big Data

Caso de uso: análisis y control de la contaminación ambiental para la gestión del tráfico y la movilidad urbana

1. Objetivos

- a) Recoger, procesar y almacenar datos sobre calidad del aire en la ciudad de Valencia en tiempo real con el fin de apoyar decisiones operativas y estratégicas en la gestión urbana, especialmente en relación con el tráfico rodado, la movilidad sostenible y la salud pública.
- b) Desarrollar modelos de inteligencia artificial que describan y predigan el comportamiento de las variables relacionadas con la contaminación ambiental.
- c) Desarrollar una solución software para soporte a la toma de decisiones operativas y estratégicas. La solución permitirá realizar análisis de los datos almacenados en el sistema a demanda del usuario, pero también incluirá una sección de avisos que se lanzarán en tiempo real ante situaciones en las que los niveles de contaminación reales difieran significativamente de los previstos. Estos avisos proporcionarán al usuario información útil para aplicar restricciones al tráfico, activar medidas de mitigación o lanzar campañas informativas.
- d) La solución software incluirá además funcionalidad de apoyo a la toma de decisiones estratégicas mediante un entorno de simulación en el que se podrán lanzar predicciones para diferentes escenarios de contaminación atmosférica según los parámetros definidos por el usuario.

2. Posibles escenarios de aplicación

El escenario en el que se enmarca este proyecto es el de un organismo público que solicita el desarrollo de una solución que integre información fiable sobre la calidad del aire para:

- Aplicar restricciones dinámicas al tráfico en función de los niveles de contaminación atmosférica.

Descripción ampliada del problema a resolver:

El sistema permitirá visualizar mapas con los niveles de contaminación atmosférica registrados en las últimas horas y previstos para la siguiente jornada. El gestor de movilidad urbana, tras un análisis de la información disponible en el sistema (real y prevista), planificará periodos de alerta ambiental en rangos temporales variables y según unos niveles de alerta preestablecidos asociados a medidas progresivas de intervención que son conocidos por todos los agentes que participan en la gestión del tráfico y la movilidad urbana:

- **Nivel 0 – Bueno:** Calidad del aire óptima. Circulación normal y sin restricciones.
- **Nivel 1 – Moderado:** Se intensifica la vigilancia en zonas con tráfico denso. Se recomienda el uso de transporte público.

- **Nivel 2 – Alto:** Se activan avisos automáticos a gestores urbanos. Se aplica reducción de vehículos autorizados en zonas sensibles como colegios, centros históricos o áreas comerciales.
- **Nivel 3 – Muy alto:** Se recomienda el teletrabajo, se limitan los accesos a zonas céntricas y se activan protocolos de salud pública.

De forma complementaria a los periodos de alerta ambiental planificados, el gestor de movilidad urbana recibirá avisos automáticos si los niveles de contaminación en tiempo real superan el umbral previsto. Esto permitirá activar medidas de forma inmediata, como restricciones de circulación, refuerzo del transporte público o campañas informativas. El aviso irá acompañado de un mapa de zonas afectadas y de una secuencia de imágenes recientes captadas por las cámaras y procesadas por el modelo de clasificación de imágenes, el cual ha detectado una anomalía y activado el aviso. Este modelo permitirá detectar señales visuales compatibles con una situación de empeoramiento de la calidad del aire, como acumulación de tráfico, baja visibilidad o presencia de humo.

Además, el sistema permitirá registrar manualmente, por parte de operarios o sensores distribuidos, información sobre puntos críticos de acumulación de emisiones, zonas con exceso de tráfico o áreas de baja ventilación. Esta información podrá utilizarse para mejorar el modelo o lanzar simulaciones y evaluar si el sistema funciona correctamente tanto en lo que respecta a la clasificación como en la recomendación de intervención.

Por último, la solución incluirá un chatbot que permitirá consultar los procedimientos de acreditación y unidades competenciales del dominio medioambiental, contaminación, gestión del tráfico y movilidad urbana.

Mejoras o adaptaciones opcionales:

- Activar zonas de bajas emisiones de manera automatizada ante episodios de alta contaminación sin intervención del gestor del servicio.
- Desarrollar funcionalidades que permitan al usuario final (gestor de movilidad) lanzar medidas correctoras de forma proactiva, no solo reactiva, para prevenir que se llegue a situaciones de alta contaminación.
- Desarrollar interfaces visuales que se puedan integrar en paneles informativos urbanos, rotativos de televisión, o aplicaciones de movilidad.
- Incorporar una capa de interoperabilidad para que los datos consolidados sobre emisiones y episodios de contaminación puedan ser utilizados para estudios de planificación urbana y sostenibilidad o en un contexto de entrenamiento federado o cualquier otro uso del ámbito de los espacios de datos.

3. Recursos a disposición de los alumnos

3.1. Ejemplo de flujo de ingesta de datos meteorológicos en tiempo real (para adaptar a datos de calidad del aire)

Como ejemplo orientativo del componente que se encargará de obtener datos origen y guardarlos en tiempo real en un almacén de datos local, se proporciona a continuación un flujo de recogida de datos meteorológicos desarrollado con Apache NiFi. El alumnado deberá analizarlo y adaptarlo al caso de uso planteado con sobre calidad del aire.

Nombre del flujo de ejemplo: el-tiempo.net Valencia (flujo meteorológico que debe ser adaptado al contexto de calidad del aire)

Objetivo: Automatizar la recolección, consolidación y almacenamiento de datos sobre calidad del aire de la ciudad de Valencia desde una fuente pública (el-tiempo.net, utilizada aquí únicamente como ejemplo meteorológico), generando datos preparados para análisis horario.

Componentes principales:

1. **InvokeHTTP**

- **Descripción:** Realiza peticiones GET al endpoint de la API cada 5 minutos.
- **URL:** <https://www.el-tiempo.net/api/json/v2/provincias/46/municipios/46250>

2. **MergeContent**

- **Descripción:** Agrupa 12 muestras (1 hora) de datos sobre calidad del aire para su análisis conjunto.

3. **JoltTransformJSON**

- **Descripción:** Transforma el JSON para homogeneizar los datos y añadir campos como fechaLectura.
- **Transformación usada:**

```
{
  "operation": "default",
  "spec": {
    "fechaLectura": "${now():toNumber()}"
  }
}
```

4. **PutMongo / PutSQL**

- **Descripción:** Almacena los datos procesados en bases de datos para su explotación posterior.
- **MongoDB:** DB = nifi, Colección = el_tiempo_net_Valencia
- **PostgreSQL:** Tabla meteorológica (configurable)

Frecuencia de ejecución: Cada 5 minutos

Formato de entrada: JSON de la API el-tiempo.net

Formato de salida: Documento JSON enriquecido, insertado en la base de datos

Tal como se puede observar en la ficha técnica, los datos crudos (json) se almacenan en una base de datos MongoDB y en formato estructurado en una base de datos PostgreSQL en la tabla meteorológica.

Los datos incluyen variables como: temperatura (°C), humedad relativa (%), velocidad y dirección del viento (km/h), probabilidad de precipitación (%), presión atmosférica (hPa), entre otros. Cada registro se asocia a una marca temporal y una localización geográfica.

La marca temporal se almacena en formato numérico por lo que se debe convertir en formato legible ejecutando la siguiente actualización en MongoDB, que se puede incorporar mediante otro processor en Apache Nifi o ejecutar periódicamente desde una consola:

```
// Seleccionamos la base de datos.
use("nifi");

// para cambiar datos de fecha en formato entero a datos tipo fecha
db.el_tiempo_net_Valencia.updateMany(
  {},
```

```
[
  {
    $set: {
      fechaLectura: { $toDate: "$fechaLectura" }
    }
  }
]
```

Como se pretende desarrollar una solución local, se pondrá a tu disposición una fuente de datos que simule la original. La fuente local incluirá un simulador con datos históricos que se comportará como si fueran datos en tiempo real. Se proporcionarán endpoints equivalentes a los originales que se propongan con latencia configurable para poder acelerar o reducir la frecuencia con la que se generan nuevos datos.

Este flujo sirve como punto de partida para diseñar un sistema de ingesta similar para datos ambientales. En este proyecto, debes crear tu propio flujo que recoja, procese y almacene datos de calidad del aire procedentes de fuentes públicas. Se espera que identifiques el endpoint del que se recogen los datos, las variables procesadas, los campos transformados y los nombres de las bases de datos según el contexto de contaminación ambiental.

3.2. Ejemplo de modelo de clasificación

Como ejemplo de modelo de clasificación, podemos imaginar una solución desarrollada para un servicio de limpieza urbana que necesita anticipar y priorizar intervenciones en función de las condiciones meteorológicas previstas para los próximos tramos horarios.

En este caso, el sistema se basa en un modelo de aprendizaje automático entrenado con datos históricos sobre precipitaciones, humedad relativa y viento, correlacionados con las decisiones de actuación del servicio en diferentes zonas de la ciudad. El flujo general consiste en recoger los datos previstos a través de un sistema de ingesta como el descrito en el apartado anterior, procesarlos y extraer variables clave (por ejemplo, intensidad de lluvia prevista y duración del episodio), que se utilizan como entrada de un clasificador supervisado.

El modelo, que puede implementarse utilizando algoritmos como Random Forest o Regresión Logística, devuelve como salida un nivel de alerta asociado a cada zona y franja horaria:

- Nivel 0: limpieza rutinaria
- Nivel 1: limpieza reducida o solo en zonas prioritarias
- Nivel 2: postergación de tareas por lluvia intensa
- Nivel 3: cancelación total de las actividades

Este modelo se despliega como servicio, recibe los datos del sistema de predicción meteorológica, y devuelve predicciones que pueden consultarse desde el sistema central o integrarse en el dashboard del gestor.

Este ejemplo no corresponde al caso real del proyecto, pero proporciona una base metodológica para desarrollar un modelo de clasificación de niveles de alerta adaptado a datos de calidad del aire en el contexto de restricciones de circulación por contaminación.

3.3. Ejemplo de modelo diagnóstico a partir de imágenes

Como complemento al modelo basado en datos meteorológicos, se puede implementar un segundo modelo orientado al análisis de imágenes urbanas captadas por cámaras públicas distribuidas en distintos puntos de la ciudad.

En el contexto del servicio de limpieza urbana, este modelo de diagnóstico por visión artificial tiene como objetivo detectar situaciones que podrían no haber sido previstas por el modelo numérico: encharcamientos, acumulación de residuos, baja visibilidad, etc. El modelo se entrena utilizando técnicas de clasificación de imágenes, como redes convolucionales (CNN), y categorías etiquetadas previamente (por ejemplo: “normal”, “charcos”, “acumulación de hojas”, “poca visibilidad”).

El sistema recibe imágenes en tiempo real desde las cámaras, las procesa en un pipeline que incluye redimensionado, normalización y clasificación, y devuelve una predicción con un nivel de confianza. Si la predicción indica una situación anómala no coherente con lo previsto para esa franja horaria (por ejemplo, imagen con charcos cuando no estaba prevista lluvia), se genera un aviso para el gestor del servicio.

Este ejemplo sirve de referencia para construir un modelo similar que, en el contexto real del proyecto, deberá analizar imágenes urbanas en relación con la contaminación atmosférica.

3.4. Ejemplo de entorno de simulación para apoyo a la toma de decisiones estratégicas

Para facilitar la toma de decisiones estratégicas a medio y largo plazo, se puede diseñar un entorno de simulación que permita explorar diferentes escenarios meteorológicos y su impacto sobre el servicio de limpieza urbana.

Este entorno consiste en una interfaz web que permite al usuario introducir manualmente parámetros como volumen de lluvia previsto, intensidad del viento o frecuencia de los episodios. A partir de estos valores, el sistema utiliza los modelos entrenados anteriormente (por ejemplo, el clasificador de niveles de alerta) para predecir qué zonas se verán más afectadas, qué recursos serán necesarios, y en qué momentos es más probable que deban modificarse las rutas o suspenderse tareas.

La simulación se ejecuta en tiempo real y presenta los resultados mediante visualizaciones gráficas (mapas de calor, cronogramas, gráficas de cargas de trabajo) para facilitar el análisis. El entorno también permite guardar y comparar distintos escenarios, lo que resulta útil para justificar decisiones o evaluar estrategias alternativas.

Este entorno de simulación es un ejemplo extrapolado del contexto de meteorología y limpieza, pero su estructura puede adaptarse fácilmente para construir un entorno de simulación orientado a contaminación ambiental y planificación del tráfico.

3.5. Ejemplo de RAG para consulta de información relativa a acreditaciones

En algunos servicios municipales, especialmente aquellos sometidos a auditoría o control externo, es necesario acceder a información regulada sobre acreditaciones, certificaciones y normativas aplicables a su ámbito de actuación. Un ejemplo sería el uso de un sistema de búsqueda basado en Recuperación Aumentada con Generación (RAG) que permita consultar, mediante lenguaje natural, documentos relacionados con procedimientos de acreditación y unidades competenciales en el ámbito meteorológico.

Este sistema se basa en un motor semántico (por ejemplo, Elasticsearch o FAISS) para almacenar y recuperar fragmentos de documentos técnicos relevantes, combinado con un modelo LLM local (como Llama2 o Mistral 7B) que se encarga de generar respuestas a partir del contenido recuperado.

El flujo típico sería:

1. El usuario introduce una pregunta del tipo: “¿Qué competencias debe acreditar un operador del sistema de predicción de lluvias?”
2. El sistema busca en su base de datos semántica los fragmentos más relevantes.
3. El LLM genera una respuesta con base en esos fragmentos, mostrando también la fuente.

Este ejemplo corresponde al ámbito meteorológico, pero puede extrapolarse para desarrollar un sistema RAG que permita consultar información sobre acreditaciones y competencias vinculadas a la gestión de la contaminación y la movilidad urbana.

Se pondrá a tu disposición modelos LLM locales como Llama2 o Mistral 7B desplegados en servidores del centro que podrán ser consumidos a través de un api de servicios (API Rest) para que no tengas que lanzar los modelos en las máquinas de desarrollo, lo que implica utilizar modelos LLM más básicos o, en cualquier caso, reducir drásticamente los recursos hardware que no están dedicados al LLM.

También se proporcionará un motor Elasticsearch en el que se podrán almacenar y recuperar los documentos.

Este componente les permitirá experimentar con uno de los paradigmas emergentes en IA aplicada a documentos: la combinación de búsqueda semántica con generación contextualizada.

3.6. Fuentes de datos

Con respecto al bloque de orígenes estructurados, el catálogo de datos abiertos del ayuntamiento de valencia (<https://valencia.opendatasoft.com>) contiene conjuntos de datos ambientales que se actualizan con bastante frecuencia (casi tiempo real). Los profesores del ciclo hemos comprobado que a partir de esos datos se pueden generar modelos con una calidad suficiente para el caso de uso que se está proponiendo por lo que es obligatorio utilizar ese origen de datos, aunque se valorará positivamente si además de los datos del Ayuntamiento de Valencia se utilizan otras fuentes.

Con respecto al origen de tipo imagen, tenemos un centenar de cámaras repartidas por la ciudad de valencia que ofrecen en abierto y en tiempo real la señal de video las 24 horas del día, los 7 días de la semana. Se puede acceder al stream de video de cada cámara a través del visualizador web <https://camaras.valencia.es/camaras/visualizador.html>

Con respecto al dominio de lenguaje natural, la Entidad Nacional de Acreditación (ENAC) otorga acreditaciones específicas en el ámbito de la gestión medioambiental. En la web (<https://www.enac.es/>) de la ENAC podéis encontrar documentos que contienen información específica sobre los procedimientos de acreditación y las unidades competenciales.

4. Pasos de implementación

Como se ha comentado en la sección de objetivos, se deben implementar tres componentes software en la solución.

El primero será un proceso de ingesta de datos, basado en el flujo de ejemplo descrito en la sección anterior, pero adaptado para obtener información sobre la calidad del aire desde una fuente pública o simulada. El flujo deberá estar configurado para consumir datos relevantes del nuevo contexto (NO₂, PM2.5, etc.) y almacenarlos correctamente para su análisis posterior.

El segundo será un conjunto de modelos de inteligencia artificial que serán consumidos por el usuario, tal como se describe a continuación. Estos modelos estarán orientados a predecir la evolución de la contaminación ambiental en determinadas zonas y a detectar automáticamente situaciones anómalas a partir de datos en tiempo real e imágenes urbanas.

El tercero es la propia aplicación de usuario (front-end), que implementará alguno de los escenarios de aplicación descritos anteriormente. Esta aplicación deberá integrar obligatoriamente un cuadro de mandos desarrollado en Power BI, que permita realizar análisis a demanda sobre los datos almacenados en el sistema. En el cuadro de mandos se incluirán tanto los datos reales como predicciones generadas por los modelos de IA. Por ejemplo:



La aplicación deberá incluir una sección de avisos que se activarán automáticamente cuando los niveles de contaminación en tiempo real superen los previstos por el sistema. Estos avisos ofrecerán al usuario información de apoyo para aplicar restricciones, modificar la planificación del tráfico o activar campañas informativas.

Será obligatorio desarrollar un modelo capaz de clasificar imágenes en función de su contenido. Tendrás a tu disposición un conjunto de imágenes no etiquetadas que se irán entregando progresivamente durante el periodo de desarrollo del proyecto para simular una dinámica de reto tipo hackathon. Este modelo se integrará en el sistema de avisos, permitiendo detectar automáticamente, a partir de imágenes urbanas captadas por cámaras, señales visuales compatibles con una situación de empeoramiento de la calidad del aire: baja visibilidad, acumulación de tráfico, presencia de humo, etc.

En este apartado será obligatorio incluir algún tipo de aviso basado en imágenes. Por ejemplo, si el modelo indica que la situación está bajo control, pero las cámaras captan imágenes con signos visibles de polución, se deberá lanzar un aviso para actualizar los niveles de alerta y modificar las acciones programadas.

Por último, la solución debe incluir un buscador tipo RAG para responder a preguntas relacionadas con el escenario de aplicación seleccionado. Será una interfaz de usuario tipo chatbot como la que se ha desarrollado en prácticas anteriores.

5. Evaluación

Este proyecto integrador tiene un peso importante en tu evaluación. Aquí demostrarás que no solo has adquirido los conocimientos de cada módulo, sino que sabes combinarlos en una solución completa que aprovecha las ventajas de cada una de las tecnologías y herramientas presentadas, aplicadas ahora al dominio de la calidad del aire y la movilidad urbana.

Deberás grabar un video de 12 minutos en el que se describa la solución desarrollada. El contenido del video debe estar dividido en tres secciones claramente diferenciadas, por si se tuviera que recortar dependiendo de la situación.

En el primer bloque, de 3 minutos, se presentará el problema a resolver, la solución desarrollada, y por qué la inteligencia artificial y el big data permiten mejorar la gestión de la contaminación ambiental y la movilidad urbana, frente a una aproximación de ingeniería del software tradicional.

En el segundo bloque, 6 minutos, se describirá a nivel técnico la solución. Se deben especificar los flujos de ingesta de datos, los algoritmos que se han utilizado para el desarrollo de los modelos, el modelo semántico de Power BI o cualquier otro artefacto de la solución.

En el tercer bloque, 3 minutos, describe cómo ha sido tu proceso de desarrollo. Qué dificultades o problemas han ocurrido. Cómo se han resuelto. También es importante mencionar otras funcionalidades adicionales que se podrían añadir o qué partes de la aplicación deberían mejorarse para que la solución fuera más robusta y estable. Por último, será muy útil un breve comentario acerca de aspectos que consideráis mejorables en el ciclo de especialización en lo que respecta a la organización, los contenidos o cualquier otro aspecto.

Eres responsable de la totalidad del desarrollo del proyecto. En caso de trabajar en grupo o desarrollo conjunto de un proyecto más completo entre todos los alumnos, todos los

integrantes deberán participar activamente en todas las tareas del proyecto y ser capaces de defender cualquier parte de la solución durante la presentación. Durante la exposición se podrá hacer una breve ronda de preguntas para verificar la comprensión global del proyecto por parte de cada alumno.

5.1. Resumen de entregables:

- Proceso de ingesta funcional con fuente local.
- Modelos IA entrenados con dataset proporcionado.
- Aplicación front-end con dashboard, alertas y chatbot.
- Vídeo de presentación (12 minutos).

5.2. Rúbrica:

Tu proyecto se evaluará según los siguientes criterios:

CRITERIO	PORCENTAJE
Integración técnica completa de los componentes	30%
Aplicación de IA/Big Data y justificación técnica	25%
Calidad del cuadro de mandos y funcionalidad de alertas	20%
Creatividad, iniciativa y escalabilidad del sistema	15%
Claridad del vídeo y reflexión final	10%

También se valorarán:

- Calidad del código (claridad, organización, buenas prácticas).
- Uso de herramientas profesionales (entornos virtuales, control de versiones, pruebas).
- Uso justificado de tecnologías emergentes (como LLMs o RAG).

Este proyecto integrador se enmarca en el Ciclo de Especialización en Inteligencia Artificial y Big Data, y supone una excelente oportunidad para aplicar, de forma práctica y transversal, todo lo aprendido en el ciclo. Se valorará especialmente la iniciativa, la autonomía y la capacidad de encontrar soluciones creativas a retos reales.

Este proyecto es también una invitación a imaginar cómo pueden utilizarse la inteligencia artificial y el análisis de datos para mejorar nuestras ciudades.

Aprovecha esta oportunidad para demostrar todo lo que sabes y desarrollar una solución con impacto real. Puede que se convierta en el primer proyecto real de tu nueva carrera profesional.